Locating systematic reviews of test accuracy studies: how five specialist review databases measure up.

Bayliss, Susan; Davenport, Clare

DOI:
10.1017/S0266462308080537

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Bayliss, S & Davenport, C 2008, 'Locating systematic reviews of test accuracy studies: how five specialist review databases measure up.', *International Journal of Technology Assessment in Health Care*, vol. 24, no. 4, pp. 403-11. https://doi.org/10.1017/S0266462308080537

International Journal of Technology Assessment in Health Care, **24:4** (2008), 403–411.

doi:10.1017/S0266462308080537

Locating systematic reviews of test accuracy studies: How five specialist review databases measure up

Sue E. Bayliss, Clare Davenport
University of Birmingham

Objectives: The aim of this study was to examine location of systematic reviews of test accuracy in five specialist review databases: York CRD's DARE and HTA databases, Medion (University of Maastricht), C-EBLM (International Federation of Clinical Chemistry), and the ARIF in-house database (University of Birmingham).
Methods: Searches were limited to the period 1996–2006. Test accuracy reviews were located using in-house diagnostic search filters and with help from database producers where databases were not confined to test accuracy reviews. References were coded according to disease area, review purpose, and test application. Ease of use, volume, overlap, and content of databases was noted.
Results: A large degree of overlap existed between databases. Medion contained the largest number ($n = 672$) and the largest number of unique ($n = 328$) test accuracy references. A combination of three databases identified only 76% of test reviews. All databases were rated as easy to search but varied with respect to timeliness and compatibility with reference management software. Most reviews evaluated test accuracy (85%) but the HTA database had a larger proportion of cost-effectiveness and screening reviews and C-EBLM more reviews addressing early test development. Most reviews were conducted in secondary care settings.
Conclusions: Specialist review databases offer an essential addition to general bibliographic databases where application of diagnostic method filters can compromise search sensitivity. Important differences exist between databases in terms of ease of use and content. Our findings raise the question whether the current balance of research setting, in particular the predominance of research on tests used in secondary care, matches the needs of decision makers.

Keywords: Information storage and retrieval; Databases, bibliographic; Technology assessment; Systematic reviews; Tests, diagnostic

There has been a growth in the volume of primary research concerned with testing over recent years, reflected in the number of systematic reviews in the area. The analysis of all systematic reviews indexed in MEDLINE in 2004 by Moher et al. (15) estimated 2,500 systematic reviews were being published annually. Although systematic reviews concerned with test performance represent a small proportion of reviews overall (approximately 8 percent) (15), their number has increased substantially over the past decade (7;12;16) substantiated by analysis of the five databases in our own study (see Supplementary Figure 1, which can be viewed online at www.journals.cambridge.org/thc).

Systematic reviews are an important resource for summarizing existing knowledge about test accuracy from

Many thanks to Julie Glanville (CRD University of York), Joseph Watine (IFCC C-EBLM), and Berna Schouten, MEDION for their help during the conduct of the research; and to Julie Glanville and Kath Wright (CRD), Joseph Watine (IFCC), Chris Hyde, and Anne Fry-Smith (University of Birmingham) for comments on earlier drafts. Source of funding: None.

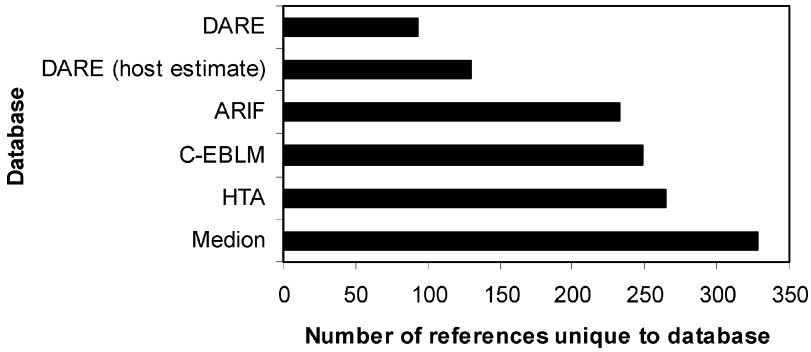

Figure 1. Unique References According to Database.

primary studies either for the busy practitioner trying to find a review to help answer a clinical question or for the reviewer wanting to assess the available evidence about a test. Databases of systematic reviews of test accuracy are also an important resource when undertaking methodological research as they can provide a representative or selective sample of reviews.

However, it is well-documented that using methodological search filters with general bibliographic databases to locate studies of test accuracy is at best unreliable (10;14;17). This would suggest that specialist review databases and in particular those devoted to reviews of test accuracy provide an alternative and possibly more reliable means of efficiently accessing these studies.

AIMS AND OBJECTIVES

To evaluate how five specialist systematic review databases perform with respect to different research requirements concerned with test accuracy in terms of (i) Overlap between databases, (ii) Utility (flexibility of searching, accessibility, compatibility with reference management software) and currency, (iii) Epidemiology of systematic reviews contained in the databases.

METHODS

Our objective was to establish as large a repository of systematic reviews of test accuracy as possible for the purposes of a piece of methodological research. This offered the opportunity to examine the databases from which the reviews were sourced in detail and to describe the epidemiology of reviews contained in these databases.

The five chosen databases all claim to contain systematic reviews as opposed to narrative reviews and commentaries. Systematic reviews have been defined as reviews that use "systematic and explicit methods to identify, select and critically appraise relevant primary research, and to extract and analyze data from studies that are included in the review" (9). The extent to which reviews contained in included databases met these criteria was not ascertained.

The specialist reviews databases included were the following (date of inception in brackets): Health technology assessment (HTA) database by means of Cochrane Library (1998) (2); Database of Abstracts of Reviews of Effects (DARE) by means of Cochrane Library (1994) (2); Medion database of diagnostic reviews (University of Maastricht) (1994) (13); International Federation of Clinical Chemists Committee of Evidence Based Laboratory Medicine (IFCC C-EBLM) reviews database (established 1996 as personal Web site of Wytze Oosterhuis, publicly available on the IFCC Web site in 2004) (1); ARIF (Aggressive Research Intelligence Facility, University of Birmingham) in-house reviews database (1996).

Of these specialist review databases, Medion and C-EBLM are devoted solely to systematic reviews of tests, whereas only a percentage of the other much larger general systematic reviews databases comprise reviews of tests (around 6 percent of HTA and 10–11 percent of DARE and ARIF). Despite the fact that at the time the research was conducted the ARIF in-house database was not accessible to the public it was chosen for inclusion on the basis that its characteristics are well known to the authors and, therefore, could be used as a point of reference for considering the other databases.

Optimal search strategies such as those created by Haynes and Wilczynski (6) for use with general bibliographic databases do not exist for use with DARE and HTA so a pragmatic filter had to be created to retrieve as many test accuracy reviews as possible, given that these reviews are not denoted by any special indexing on these two databases. The ARIF database tags diagnostic and screening reviews as such on inclusion so these were easily retrieved.

Searches for all systematic reviews of test accuracy in each database were carried out in January 2007 for the period 1996–2006. Ease of use of databases (see Table 1) was noted. Given the number of references likely to be retrieved and the difficulties caused by poor indexing of systematic reviews of test accuracy already described, searches of HTA and DARE

Table 1. Finding Systematic Reviews of Test Accuracy: Comparing Features of Specialist Review Databases

	Medion	DARE	HTA	IFCC's C-EBLM	ARIF
Owned by	University of Maastricht	University of York CRD	INAHTA/ University of York CRD	International Federation of Clinical Chemistry C-EBLM Committee	ARIF University of Birmingham
Document source	Majority sourced from MEDLINE, some from other sources	Handsearching, scanning databases(list on Web site)	Scanning Web sites, INAHTA project submissions every 6 months	Searching MEDLINE and other resources using filters, contact with experts, HTA sites	Scanning and alerting services
Study type	Systematic reviews. Separate methodology and genetics databases	Systematic reviews	Systematic reviews, other types of health technology assessment	Systematic reviews	Systematic reviews. Separate methodology database
Overall size of databases (Jan 2007)	1500 0verall: 796 – reviews, 597 – methods, 119- genetics	4539	6175	555	8670
Dataset	Au, ti, source	Au, ti, source	Au, ti, source	Au, ti, source	Au, ti, source
Abstract	√	√	√	√	×
Textword search	√	√	√	×	√
Indexed by	Keywords IPCP codes	MeSH	MeSH	Keywords	Keywords
Links to text	×	√	√	×	×
Ease of use	√	√	√	√	√
Advanced search	√	√	√	×	×
Quality assured	×	√	×	×	×
Updating	Periodic	Monthly	Monthly	Twice per year	Weekly
Gives no of hits	√	√	√	√	√
Compatible for reference management software	x	√	√	×	√
Sort facility	√	√	√	√	×
Disadvantages	Web site is sometimes inaccessible because of technical problems No help facility	Cannot search for test accuracy reviews separately at present	Test accuracy studies not identified separately as such	Database not easy to locate from IFCC Web site. Potential language bias (English reviews only)	Currently only available in-house (plans to make it accessible via ARIF Web site).
Advantages	First established specialist database of reviews of test accuracy. Most unique references	Special index field "Reference standard against which new test was compared" to allow for detailed indexing of reviews of test accuracy studies.	International focus. Good resource for reviews concerned with screening	Special focus on laboratory medicine Well indexed. Lot of unique references	Most current as updated on weekly basis Separate subset of diagnostic reviews easy to locate

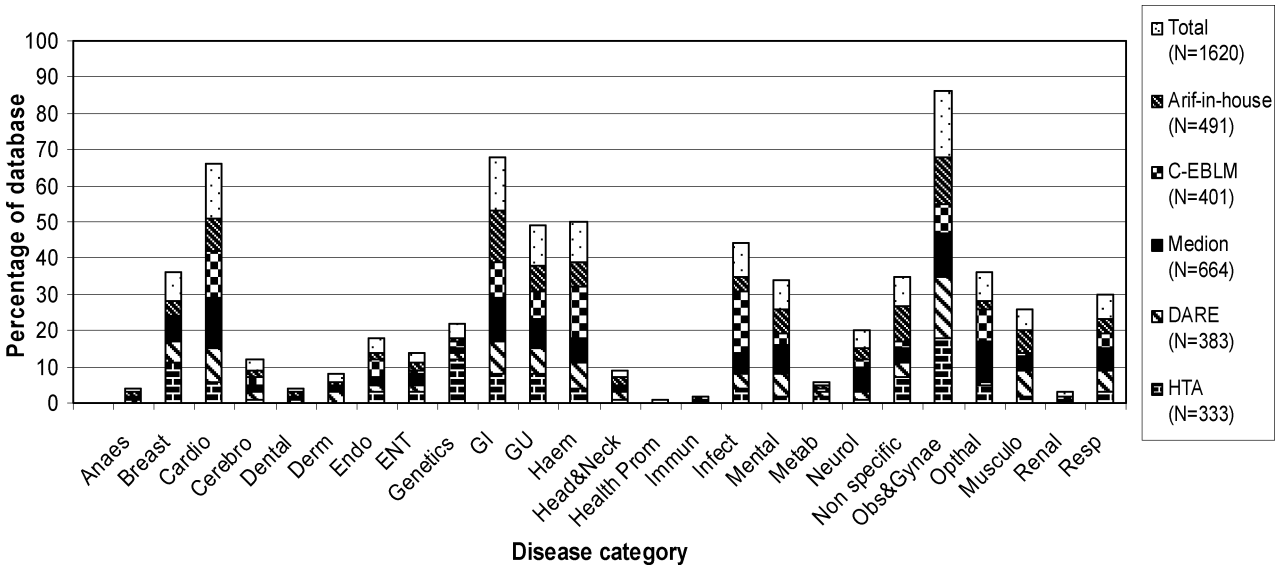

Figure 2. Percentage of each database accounted for by disease category.

were limited to MeSH index terms to make them as specific as possible. To select the most appropriate MeSH terms for our filter an analysis by Leeflang et al. (10) which compared the performance of twelve validated diagnostic search filters was consulted. We selected the most frequently used MeSH term in these filters: "Sensitivity and Specificity" (exp) (92 percent of filters) to which the term "Mass Screening" was added to capture a variety of testing applications. The MeSH term "Diagnosis" (exp) or text word "diagnostic" greatly reduced the specificity of the searches and so were not used. We sought to verify the performance of our filter with the help of in-house searches of DARE and HTA performed by database producers CRD (Centre for Reviews and Dissemination, University of York). Diagnostic reviews in DARE are coded in-house but searching in this way is not currently available on the general database interface. A search of the HTA database was undertaken by database producers using their preferred terms. The ARIF database was searched on the subset *diagnosis* as well as the text word *screening* and false positive hits were identified by scrutiny of retrieved records (see Supplementary Table 1, which can be viewed at www.journals.cambridge.org/thc, for terms used in filters).

Scrutiny of retrieved records for false positive hits (reviews not concerned with diagnosis) also allowed investigation of the specificity of the filter (see flow of references Supplementary Figure 2, which can be viewed at www.journals.cambridge.org/thc). All records for the relevant period in the specialist diagnostic reviews databases, Medion and C-EBLM, were included. Reference Manager v 11 for Windows was used to store downloaded records from

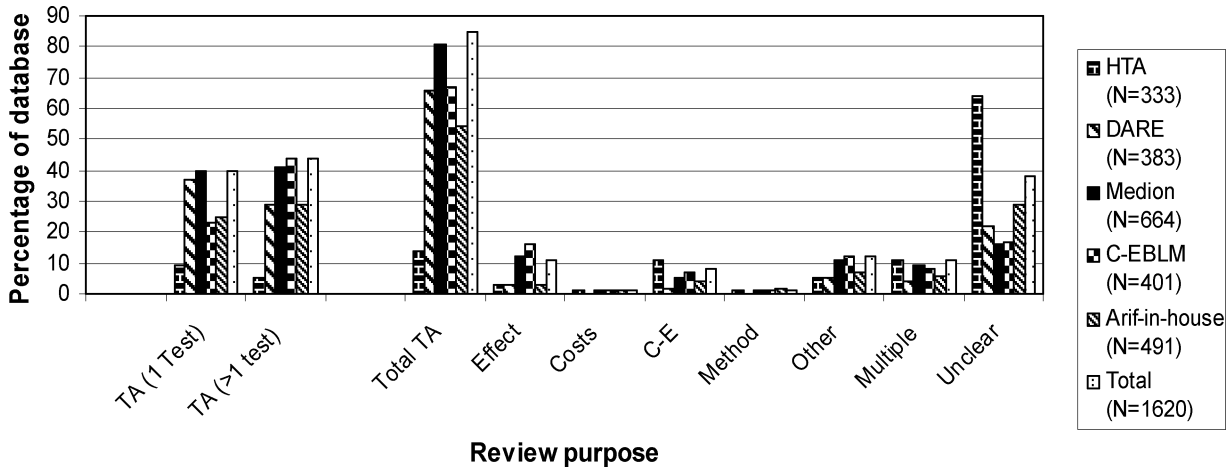

Figure 3. Percentage of each database according to review purpose.

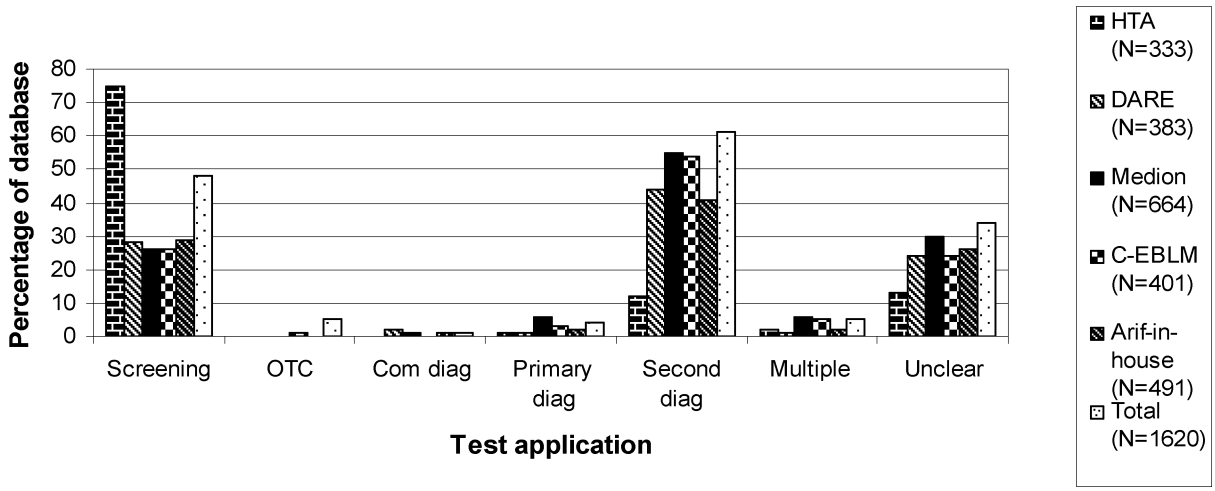

Figure 4. Percenntage of individual databases according to test application.

DARE, HTA and ARIF. C-EBLM and Medion are not compatible with Reference Manager at this time and citations from these databases were, therefore, added manually.

Exclusion Criteria

The focus for the methodological review was systematic reviews concerned with assessment of test accuracy, either in isolation or as part of a broader evaluation of tests. Thus our search aimed to identify systematic reviews concerned with evaluation of test accuracy, test effectiveness, test cost-effectiveness, and test accuracy methodological reviews. Papers were excluded only if they were not concerned with test evaluation or they were not reviews.

Coding of References

References were tagged according to their database source. In addition epidemiological characteristics of test accuracy reviews were noted based on review title. The disease topic area or areas the review was addressing was noted. To ensure consistency a pro-forma was used as in some instances a topic could be placed in more than one category. For example tuberculosis was consistently placed in "infectious diseases" rather than "respiratory". Further detail of the method of classification is available from the authors on request.

The purpose of the review was coded as "test accuracy" only, "costs" of testing, "effectiveness" of testing, "cost-effectiveness" of testing, "methodological" reviews or "other", (concerned with test acceptability; descriptive accounts of promising disease markers, tests and test strategies; consequences of tests error; organization of testing programs; morphological studies; methods of test execution). Test accuracy reviews were further subdivided into those concerned with estimation of test accuracy of single tests or with estimation of accuracy of more than one test.

The clinical setting in which the test was being evaluated for use was noted. Test setting was defined as the likely origin of patients to be tested and not the setting in which the test was to be applied. Thus for example ultrasound examination and X-rays could be initiated and acted on in primary care but the tests themselves would be likely to take place in a secondary care setting. Reviews were coded as being concerned with tests to be used in a screening context (encompassing population based and targeted screening programs), over the counter, in the community, primary care, secondary care or for use in multiple settings.

The search facility in Reference Manager was used to identify yield of references by single database and database combinations and to map characteristics of test accuracy reviews contained in the databases.

RESULTS

Performance of Pragmatic Search Filter in General Specialist Review Databases

The flow of references is illustrated in Supplementary Figure 1 (www.journals.cambridge.org/thc). For those databases with MeSH search facilities (DARE and HTA) the pragmatic filter performed variably for identification of reviews concerned with testing. There were 89 false positive hits for DARE (19 percent of DARE hits) and 9 in the HTA database (3 percent of HTA hits). The number of false positives generated by searching the ARIF in-house database using the terms diagnosis and screening was low; $n = 13$ (3 percent). In the HTA database only 16 (5 percent) of hits were not reviews. In the ARIF database 2 (<1 percent) of hits had been wrongly added to the database as reviews when in fact they were primary research (mostly case series).

Our pragmatic filter on the DARE database identified 72 percent of test accuracy reviews identified by DARE

producers. Our search yielded 383 relevant references for period 1996–2006 of which 24 percent (93) were unique to DARE. The DARE in-house search yielded 542 tagged references for the same period and we, therefore, estimated 130 of these references would be unique to DARE (24 percent of 542). Both estimates are important; our estimate of 383 is likely to approximate to the yield from a search of DARE on the public interface whilst the database producer estimate is a more valid representation of the number of test accuracy reviews within DARE. Reviews identified by in-house searching of the HTA database yielded fewer hits than searches using our pragmatic filter (HTA in-house search $n =$ 172 and our search $n = 333$ between 1996 and 2006). For the purposes of calculating yield of relevant references for single databases and across multiple databases we, therefore, used our estimate of 333.

Duplication across Databases

Yield of Test Accuracy Reviews by Single Databases. One record from Medion (a letter) had been erroneously included in the database. All of the records in the C-EBLM databases were reviews. Both the Medion database and the C-EBLM database contained references not concerned with evaluation of test accuracy (1 percent of Medion records and 9 percent of C-EBLM records), all investigating putative causal associations between laboratory-based markers and disease.

If searches were restricted to a single database and where necessary use of a pragmatic search filter, the largest number of test accuracy reviews would be identified by using Medion, followed by the ARIF in-house database, C-EBLM, DARE and finally HTA (see Supplementary Figure 2, at www.journals.cambridge.org/thc). Using the DARE database producer estimate revises the proportion of reviews that would be identified by DARE to 30 percent.

Unique References by Database. After removing reviews not concerned with diagnosis and primary research papers Medion had the most unique test accuracy review references (references not contained in any other database) ($n = 328$) followed by the HTA database ($n = 264$), C-EBLM database ($n = 248$) and the ARIF database ($n = 232$). DARE had the least number of unique test accuracy review references ($n = 93$). Using the DARE database producer estimate would increase the number of references unique to the DARE database to 130 but would not change its rank order.

Yield of Test Accuracy Reviews from Searches of Database Combinations. Supplementary Table 2 online (www.journals.cambridge.org/thc) documents the yield of reviews for combinations of 2 and 3 databases. A combination of three publicly available databases (C-EBLM, Medion and HTA) at best yielded 1,232 references (76 percent of the total). A combination of Medion and the HTA database or Medion and C-EBLM yielded 948 and 952 references (59 percent of the total). The lowest yield of references was obtained by a combination of the DARE and HTA databases (561; 35 percent) but it must be noted that this low yield may be explained by the fact that DARE is a selective, quality assured resource.

Content of Databases: Characteristics of Indexed Test Accuracy Reviews

This analysis is based on the content of the 1,620 test accuracy reviews identified by our search filter in DARE and HTA, those appropriately tagged in the ARIF database and all reviews contained in Medion and C-EBLM (see Supplementary Figure 2, at www.journals.cambridge.org/thc) for the period 1996–2006. Description of review characteristics was based on review title.

Disease Topic Area. Testing in the disease area of obstetrics and gynecology accounts for between 8 percent and 18 percent, median 13 percent of citations (18 percent overall). Cardiovascular disease ("cardio") and gastrointestinal disease ("GI") were also prominent accounting for between 9 percent and 15 percent of citations, (15 percent overall). In addition ophthalmology was prominent in the Medion database (11 percent citations). The high proportion of reviews concerned with infectious disease ("Infec") and hematology ("Haem") in the C-EBLM database is probably a reflection of the laboratory emphasis of this database. The relatively high proportion of genetic testing reviews in the HTA database (12 percent) may be a reflection of the priority of this topic area as an emerging health technology (19). It should be noted that the Medion database has a separate section devoted solely to reviews concerned with genetic testing which was not included in our analysis. The number of reviews in the Medion genetics section over our period of study was 119. This would increase the proportion of genetics reviews in Medion to be close to 20 percent of the total across the general reviews and genetics sections, compared with the 1 percent indicated in Supplementary Figure 2 (www.journals.cambridge.org/thc).

Review Purpose. With the exception of the HTA database most reviews were concerned solely with the estimation of test accuracy (66–81 percent across 4 databases; 85 percent of all citations). This is not an unexpected finding as in the context of health technology assessment evaluation of the acceptability, effectiveness and cost-effectiveness of tests is important. Indeed the HTA database had the highest proportion of reviews concerned with the evaluation of cost-effectiveness of tests (11 percent). Our classification did not allow discrimination between accuracy being evaluated at different stages of test development. It was evident from scrutiny of titles and abstracts that the C-EBLM database contained a larger proportion of reviews concerned with early test development using a case control design. This was in contrast to other databases where the predominant type of

test accuracy study was in a clinical setting, (screening, diagnosis, prognosis or disease monitoring) (5;18).

Relatively few test accuracy reviews were concerned with evaluation of effectiveness alone, costs alone, or as part of the review process highlighted an area of methodology. The proportion of reviews for which the purpose was unclear was high (16–64 percent across 5 databases and 38 percent overall) and it is unknown what impact the accurate coding of this subset would have on the distribution of review purpose across databases.

Clinical Setting in which Tests Are Applied. There was a striking preponderance of tests evaluated in secondary care and screening contexts across all databases. Overall only 4 percent of reviews evaluated tests for use in primary care (1–6 percent across individual databases). Secondary care and screening would still dominate as research settings even if all of the reviews coded as "unclear setting" were in fact evaluations of tests in primary care.

Features of Databases

A comparison of the different features of the five databases is presented in Table 1. Whereas the ARIF database is least sophisticated in terms of searching and content of individual records, it is the most up-to-date of the five and test accuracy reviews are easily retrieved as they are tagged on inclusion. However, although there are plans to make it accessible by means of the ARIF website, currently this database is not accessible to the public. DARE and HTA offer a much more sophisticated product in terms of search and retrieval but test accuracy reviews are not coded separately. DARE is the only database to contain abstracts of reviews that have been quality-assessed, containing a summary of the review together with a critical commentary about its overall quality and as a result of this is a selective rather than a comprehensive resource. Medion is the longest established specialist database devoted to test accuracy reviews and is user-friendly as well as comprehensive with separate smaller databases devoted solely to reviews of genetic tests and methodological papers. The IFCC's C-EBLM database differs quite markedly in content from the other databases with its emphasis on laboratory tests and is a well indexed supplementary resource, although it is updated only twice per year.

FUTURE DEVELOPMENTS

The Cochrane Diagnostic Test Register Group (DTRG) aims to *"develop a clean and comprehensive register of reports of diagnostic test accuracy studies"* (3;4). Their register will contain primary studies concerned with test accuracy in screening or diagnostic contexts and will be the equivalent of Cochrane's CENTRAL trials register. The DTRG will also support review groups in the production of diagnostic reviews, the first pilot review is likely to be published in 2008. The reviews will be added to the Cochrane Database of Systematic Reviews (CDSR) and flagged as diagnostic reviews. The development of the register is currently in its early stages and although it will lead to improved review methodology, the number of reviews will grow slowly. In the meantime those wishing to access up to date and comprehensive secondary research evidence concerning testing will need to rely on the sources outlined above or apply methodological filters for retrieving test accuracy reviews to bibliographic databases such as MEDLINE.

SUMMARY

Specialist review databases are an important resource in terms of providing an easily accessible and reliable means of locating the increasing number of systematic reviews concerned with testing.

Specialist review databases that are not confined to reviews of test accuracy (DARE, HTA, ARIF) may be searched quite easily for a specific type of test but when a large number of reviews of any type of test are required for methodological purposes then some sort of search filter must be applied. This research was concerned with identifying as large a number of systematic reviews of test accuracy in DARE and HTA as possible in a short space of time. Our experience suggests that one size does not fit all and in particular a range of terms reflecting a variety of potential applications of tests (for example screening and diagnosis) and test accuracy outcomes in common use, such as sensitivity and specificity, should be used.

There is a large degree of overlap between databases which is an important finding for those conducting research where a comprehensive sample of reviews is not the aim. The Medion database yielded the largest number of unique references and appeared to contain a representative sample of reviews in terms of disease area, test application and review purpose. However, timeliness and reliability are potential disadvantages of the Medion database. The DARE database yielded the lowest number of unique references but has the advantage of full abstracts and advanced search facilities. The DARE database is also quality assured. However, the quality of systematic reviews of test accuracy is generally poor and is an area in development (16), therefore, the extent to which quality assessment might compromise yield is unknown.

With respect to review purpose, the C-EBLM database appears to focus on reviews concerned with early test development whilst HTA and DARE have a larger proportion of reviews concerned with evaluating the effectiveness of tests. The HTA database also has a relatively high proportion of reviews concerned with the evaluation of the cost-effectiveness of tests. NHS EED which contains over 50,000 abstracts of quality assessed economic evaluations is another obvious resource for reviewers concerned with cost-effectiveness but was not included in this analysis (2).

With respect to test application, all databases had a high proportion of reviews concerned with tests applied in

a screening context, in particular the HTA database. This is not a surprising finding when one considers the costs and risks associated with population-based screening and the existence of well developed systems for evaluating screening programs. However, evaluation of tests for use in diagnosis in secondary care also predominated across all databases. This may be a reflection of the relatively higher unit cost of many tests used in secondary care. However, the volume of tests conducted in primary care is high and continues to increase (11), probably in part facilitated by the move of chronic disease monitoring from secondary to primary care (8). Our research suggests that there is a need to examine the balance of test accuracy research being conducted across care settings.

With respect to the quality of reviews contained in databases DARE is the only database that is quality assured. Whilst this is often an advantage it may compromise generalizability, for example where the purpose of research is to reflect current practice.

For researchers requiring comprehensiveness a combination of three databases at best identified between 69 percent and 76 percent of citations (see Supplementary Table 2, available online at www.journals.cambridge.org/thc). In this context it is important to note that at present Medion and C-EBLM cannot be downloaded into reference management software which has implications for workload.

An algorithm suggesting suitability of databases for different search requirements may be viewed online as Supplementary Table 3 (www.journals.cambridge.org/thc).

LIMITATIONS OF THE STUDY AND IMPLICATIONS FOR FURTHER RESEARCH

Our pragmatic filters (see Supplementary Table 1, available online at www.journals.cambridge.org/thc) may have missed relevant citations in databases where the content was not solely concerned with testing. The main impact of any omissions would be to underestimate the contribution of databases in terms of yield in our analysis although we cannot rule out the possibility that a basic search would skew results toward references of a certain content. The pragmatic filter appears to have performed well in the HTA database although further research would be needed to verify its performance compared with other search strategies. It did not perform so well in the DARE database and until coded access to test accuracy reviews is made possible on the public interface of DARE, a more sophisticated search strategy than the one we adopted should probably be advocated.

For pragmatic reasons we coded review characteristics based on the review title alone and as a result errors in classification may have occurred; many reviews were coded as having an unclear setting or review purpose. However, this is an important finding in itself. Further research is needed to identify whether lack of clarity in review titles is a reflection of lack of clarity in review methods and/or reporting.

The purpose of our research was to identify reviews concerned in whole or in part with test accuracy. It is likely that reviews concerned with any type of test evaluation would include the terms sensitivity, specificity or screening but our search strategy may have missed reviews where the focus was on test costs, test effectiveness and test cost-effectiveness. Our analysis did not include the NHS EED database or the CDSR database. CDSR does not claim to include systematic reviews of test accuracy studies although our search filter identified 16 relevant hits from CDSR between 1996 and 2006 concerned with various aspects of screening. Using our filter in NHS EED between 1996 and 2006 identified in excess of 800 hits. Without further research we cannot comment on the relevance of these citations or their content, but NHS EED represents an important resource for those interested in cost-effectiveness issues applied to testing (2).

CONCLUSIONS

There is a large and increasing number of reviews of test accuracy. Given the widely held concern that applying methodological search filters to capture test accuracy research does not provide adequate sensitivity for systematic review purposes, specialist review databases are an important resource for identifying such reviews, although comprehensiveness is likely to require a search across multiple databases. Medion and C-EBLM offer an efficient option for researchers and reviewers. Important differences between the specialist reviews databases should be borne in mind when choosing a resource and dependent on the purpose of research. Reviews contained in the specialist databases scrutinized are largely concerned with estimation of test accuracy and with application of tests in secondary care or for screening at present. There is a paucity of reviews concerned with the application of tests for diagnosis in primary care.

Resources such as these can change rapidly so a watching brief is important when searching this area. Since writing this article, two developments of note have taken place. From the end of July 2008 the ARIF database will be accessible to the public by means of the ARIF Web site at http://www.arif.bham.ac.uk/. Regrettably the C-EBLM database is no longer accessible by means of the IFCC Web site as there is uncertainty over its future, but at present those, who are interested can request a copy of the database in Excel format from Joseph Watine (j.watine@ch-rodez.fr).

CONTACT INFORMATION

Susan E Bayliss, BA (s.bayliss@bham.ac.uk), Information Specialist, **Clare Davenport**, BSc, MBChB, MSc (c.f.davenport@bham.ac.uk), Clinical Research Fellow, Department of Public Health and Epidemiology, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

REFERENCES

1. C-EBLM Database of Systematic Reviews and Meta-Analyses in Laboratory Medicine. http://www.fescc.org/divisions/emd/c-eblm/search.asp?id=1.
2. Cochrane Library (incorporating DARE HTA CDSR and EED databases). http://www3.interscience.wiley.com/cgi-bin/mrwhome/106568753/HOME?CRETRY=1&SRETRY=0.
3. Cochrane Diagnostic Test Accuracy Working Group. http://srdta.cochrane.org/en/index.html.
4. Cochrane Screening and Diagnostic Test Methods Group. http://www.cochrane.org/docs/sadtdoc1.htm.
5. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making*. 1991;11:88-94.
6. Haynes RB, Wilczynski NL Optimal search strategies for retrieving scientifically strong studies of diagnosis from Medline: Analytical survey *BMJ.* 2004;328:1040.
7. Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy. *J Clin Epidemiol*. 1995;48:119-130.
8. Kaag ME, Wijkel D, de Jong D. Primary health care replacing hospital care – the effect on quality of care. *Int J Qual Health Care*. 1996;8:367-373.
9. Khan K, ter Riet G, Glanville J, Sowden A, Kleijnen J, eds. *Undertaking systematic reviews of research on effectiveness. CRD's guidance for those carrying out or commissioning reviews*. CRD report Number 4. 2nd ed. NHS Centre for reviews and Dissemination, University of York; March 2001. http://www.york.ac.uk/inst/crd/pdf/crd4_content.pdf. Accessed 22 February 2008.
10. Leeflang MMG, Scholten RJPM, Rutje AWS, Reitsema JB, Bossuyt PMM. Use of methodological search filters to identify diagnostic accuracy studies can lead to the omission of relevant studies. *J Clin Epidemiol*. 2006;59:234-240.
11. Leurquin P, Van Casteren V, De Maeseneer J. Use of blood tests in general practice: A collaborative study in eight European countries. Eurosentinel Study Group. *Br J Gen Pract.* 1995;45:21-25.
12. Lijmer JG, Mol BW, Heisterkamp SH et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA*. 1999;282:1061-1066.
13. Medion. http://www.mediondatabase.nl/.
14. Mitchell R, Rinaldi F, Craig J. *Performance of published search strategies for studies of diagnostic test accuracy (SDTAs) in MEDLINE and EMBASE*. XIII Cochrane Colloquium 2005, October 22–26, Melbourne, Australia.
15. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med*. 2007;4:e78.
16. Oosterhuis WV, Niessen RW, Bossuyt PM. The science of systematic reviews of diagnostic tests. *Clin Chem Lab Med*. 2000;38:577-588.
17. Ritchie G, Glanville T, Lefebvre C. Do published search filters to identify diagnostic test accuracy studies perform adequately *Health Info Libr J*. 2007;24:188-192.
18. Sackett DL, Haynes RB. The architecture of diagnostic research. In: Knottnerus A, ed. *The evidence base of clinical diagnosis*. London: BMJ Books; 2002:19-38.
19. Storz P, Kolpatzik K, Perleth M, Klein S, Haussler B. Future relevance of genetic testing: A systematic horizon scanning analysis. *Int J Technol Assess Health Care*. 2006;23:495-504.